Análisis de Datos (Titanic)

Paso 1 : Carga de Datos

```matlab
% Set up the Import Options and import the data
opts2 = delimitedTextImportOptions("NumVariables", 12);

% Specify range and delimiter
opts2.DataLines = [2, Inf];
opts2.Delimiter = ",";

% Specify column names and types
opts2.VariableNames = ["PassengerId", "Survived", "Pclass", "Name", "Sex", "Age", "SibSp", "Parch", "Ticket", "Fare", "Cabin", "Embark
opts2.VariableTypes = ["double", "double", "double", "string", "categorical", "double", "double", "double", "double", "double", "strin

% Specify file level properties
opts2.ExtraColumnsRule = "ignore";
opts2.EmptyLineRule = "read";

% Specify variable properties
opts2 = setvaropts(opts2, ["Name", "Cabin"], "WhitespaceRule", "preserve");
opts2 = setvaropts(opts2, ["Name", "Sex", "Cabin", "Embarked"], "EmptyFieldRule", "auto");

% Import the data
titanic = readtable("C:\Users\MarcoA\Documents\MATLAB\u_sabana\titanic_dataset.csv", opts2);

% Clear temporary variables
clear opts2

% Display results
titanic
```

titanic = 891×12 table

	PassengerId	Survived	Pclass	Name	Sex	Age	S
1	1	0	3	"Braund, Mr. Owen Harris"	male	22	
2	2	1	1	"Cumings, Mrs. John Bradley (Florence Briggs Thayer)"	female	38	
3	3	1	3	"Heikkinen, Miss. Laina"	female	26	
4	4	1	1	"Futrelle, Mrs. Jacques Heath (Lily May Peel)"	female	35	
5	5	0	3	"Allen, Mr. William Henry"	male	35	
6	6	0	3	"Moran, Mr. James"	male	NaN	
7	7	0	1	"McCarthy, Mr. Timothy J"	male	54	
8	8	0	3	"Palsson, Master. Gosta Leonard"	male	2	
9	9	1	3	"Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)"	female	27	
10	10	1	2	"Nasser, Mrs. Nicholas (Adele Achem)"	female	14	
11	11	1	3	"Sandstrom, Miss. Marguerite Rut"	female	4	
12	12	1	1	"Bonnell, Miss. Elizabeth"	female	58	
13	13	0	3	"Saundercock, Mr. William Henry"	male	20	
14	14	0	3	"Andersson, Mr. Anders Johan"	male	39	
15	15	0	3	"Vestrom, Miss. Hulda Amanda Adolfina"	female	14	
16	16	1	2	"Hewlett, Mrs. (Mary D Kingcome) "	female	55	
17	17	0	3	"Rice, Master. Eugene"	male	2	
18	18	1	2	"Williams, Mr. Charles Eugene"	male	NaN	
19	19	0	3	"Vander Planke, Mrs. Julius (Emelia Maria Vandemoortele)"	female	31	
20	20	1	3	"Masselmani, Mrs. Fatima"	female	NaN	
21	21	0	2	"Fynney, Mr. Joseph J"	male	35	
22	22	1	2	"Beesley, Mr. Lawrence"	male	34	
23	23	1	3	"McGowan, Miss. Anna "Annie""	female	15	
24	24	1	1	"Sloper, Mr. William Thompson"	male	28	
25	25	0	3	"Palsson, Miss. Torborg Danira"	female	8	
26	26	1	3	"Asplund, Mrs. Carl Oscar (Selma Augusta Emilia Johansson)"	female	38	
27	27	0	3	"Emir, Mr. Farred Chehab"	male	NaN	
28	28	0	1	"Fortune, Mr. Charles Alexander"	male	19	
29	29	1	3	"O'Dwyer, Miss. Ellen "Nellie""	female	NaN	
30	30	0	3	"Todoroff, Mr. Lalio"	male	NaN	
31	31	0	1	"Uruchurtu, Don. Manuel E"	male	40	
32	32	1	1	"Spencer, Mrs. William Augustus (Marie Eugenie)"	female	NaN	
33	33	1	3	"Glynn, Miss. Mary Agatha"	female	NaN	
34	34	0	2	"Wheadon, Mr. Edward H"	male	66	
35	35	0	1	"Meyer, Mr. Edgar Joseph"	male	28	
36	36	0	1	"Holverson, Mr. Alexander Oskar"	male	42	
37	37	1	3	"Mamee, Mr. Hanna"	male	NaN	
38	38	0	3	"Cann, Mr. Ernest Charles"	male	21	
39	39	0	3	"Vander Planke, Miss. Augusta Maria"	female	18	
40	40	1	3	"Nicola-Yarred, Miss. Jamila"	female	14	
41	41	0	3	"Ahlin, Mrs. Johan (Johanna Persdotter Larsson)"	female	40	
42	42	0	2	"Turpin, Mrs. William John Robert (Dorothy Ann Wonnacott)"	female	27	

43	43	0	3	"Kraeff, Mr. Theodor"	male	NaN
44	44	1	2	"Laroche, Miss. Simonne Marie Anne Andree"	female	3
45	45	1	3	"Devaney, Miss. Margaret Delia"	female	19
46	46	0	3	"Rogers, Mr. William John"	male	NaN
47	47	0	3	"Lennon, Mr. Denis"	male	NaN
48	48	1	3	"O'Driscoll, Miss. Bridget"	female	NaN
49	49	0	3	"Samaan, Mr. Youssef"	male	NaN
50	50	0	3	"Arnold-Franchi, Mrs. Josef (Josefine Franchi)"	female	18
51	51	0	3	"Panula, Master. Juha Niilo"	male	7
52	52	0	3	"Nosworthy, Mr. Richard Cater"	male	21
53	53	1	1	"Harper, Mrs. Henry Sleeper (Myna Haxtun)"	female	49
54	54	1	2	"Faunthorpe, Mrs. Lizzie (Elizabeth Anne Wilkinson)"	female	29
55	55	0	1	"Ostby, Mr. Engelhart Cornelius"	male	65
56	56	1	1	"Woolner, Mr. Hugh"	male	NaN
57	57	1	2	"Rugg, Miss. Emily"	female	21
58	58	0	3	"Novel, Mr. Mansouer"	male	28.5000
59	59	1	2	"West, Miss. Constance Mirium"	female	5
60	60	0	3	"Goodwin, Master. William Frederick"	male	11
61	61	0	3	"Sirayanian, Mr. Orsen"	male	22
62	62	1	1	"Icard, Miss. Amelie"	female	38
63	63	0	1	"Harris, Mr. Henry Birkhardt"	male	45
64	64	0	3	"Skoog, Master. Harald"	male	4
65	65	0	1	"Stewart, Mr. Albert A"	male	NaN
66	66	1	3	"Moubarek, Master. Gerios"	male	NaN
67	67	1	2	"Nye, Mrs. (Elizabeth Ramell)"	female	29
68	68	0	3	"Crease, Mr. Ernest James"	male	19
69	69	1	3	"Andersson, Miss. Erna Alexandra"	female	17
70	70	0	3	"Kink, Mr. Vincenz"	male	26
71	71	0	2	"Jenkin, Mr. Stephen Curnow"	male	32
72	72	0	3	"Goodwin, Miss. Lillian Amy"	female	16
73	73	0	2	"Hood, Mr. Ambrose Jr"	male	21
74	74	0	3	"Chronopoulos, Mr. Apostolos"	male	26
75	75	1	3	"Bing, Mr. Lee"	male	32
76	76	0	3	"Moen, Mr. Sigurd Hansen"	male	25
77	77	0	3	"Staneff, Mr. Ivan"	male	NaN
78	78	0	3	"Moutal, Mr. Rahamin Haim"	male	NaN
79	79	1	2	"Caldwell, Master. Alden Gates"	male	0.8300
80	80	1	3	"Dowdell, Miss. Elizabeth"	female	30
81	81	0	3	"Waelens, Mr. Achille"	male	22
82	82	1	3	"Sheerlinck, Mr. Jan Baptist"	male	29
83	83	1	3	"McDermott, Miss. Brigdet Delia"	female	NaN
84	84	0	1	"Carrau, Mr. Francisco M"	male	28
85	85	1	2	"Ilett, Miss. Bertha"	female	17
86	86	1	3	"Backstrom, Mrs. Karl Alfred (Maria Mathilda Gustafsson)"	female	33
87	87	0	3	"Ford, Mr. William Neal"	male	16
88	88	0	3	"Slocovski, Mr. Selman Francis"	male	NaN
89	89	1	1	"Fortune, Miss. Mabel Helen"	female	23
90	90	0	3	"Celotti, Mr. Francesco"	male	24
91	91	0	3	"Christmann, Mr. Emil"	male	29
92	92	0	3	"Andreasson, Mr. Paul Edvin"	male	20
93	93	0	1	"Chaffee, Mr. Herbert Fuller"	male	46
94	94	0	3	"Dean, Mr. Bertram Frank"	male	26
95	95	0	3	"Coxon, Mr. Daniel"	male	59
96	96	0	3	"Shorney, Mr. Charles Joseph"	male	NaN
97	97	0	1	"Goldschmidt, Mr. George B"	male	71
98	98	1	1	"Greenfield, Mr. William Bertram"	male	23
99	99	1	2	"Doling, Mrs. John T (Ada Julia Bone)"	female	34
100	100	0	2	"Kantor, Mr. Sinai"	male	34

Paso 2: Exploración inicial de los datos

```
% Ver cuántos valores faltantes hay en cada columna
missing = sum(ismissing(titanic))
```

```
missing = 1×12
    0    0    0    0    0  177    0    0  230    0    0    2
```

Paso 3 : Limpieza de los Datos

Sobre el archivo podemos ver

- Tiene un total de 12 columnas como Id del pasajero, sexo, edad, tarifa, cabina, entre otras y 891 filas
- Se tienen 3 tipos de datos Flotantes, Enteros y Strings
- Se tienen una gran cantidad de Missing Values (Nan).
- Columnas como PassengerId, Survived, Pclass, Name, Sex, SibSp, y Parch no tienen valores faltantes
- Las columnas con mayores valores faltantes pueden requeridad imputar los valores faltantes o si no sirven para el análisis excluirlos
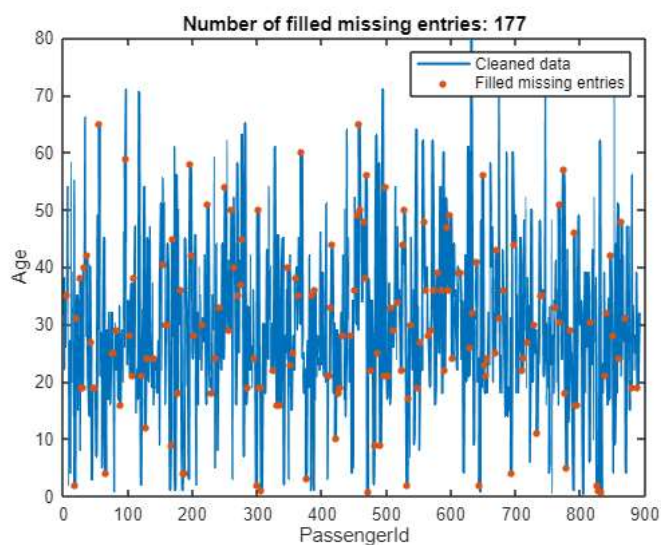
```
% Standardize missing data
newTable = standardizeMissing(titanic,"NaN");
% Fill missing data
[newTable,missingIndices2] = fillmissing(newTable,"previous", ...
    "SamplePoints",titanic.PassengerId);

% Display results
figure

% Plot cleaned data
plot(titanic.PassengerId,newTable.Age,"SeriesIndex",1,"LineWidth",1.5, ...
    "DisplayName","Cleaned data")
hold on

% Plot filled missing entries
plot(titanic.PassengerId(missingIndices2(:,6)),newTable.Age(missingIndices2(:,6)),".", ...
    "MarkerSize",12,"SeriesIndex",2,"DisplayName","Filled missing entries")
title("Number of filled missing entries: " + nnz(missingIndices2(:,6)))

hold off
legend
ylabel("Age")
xlabel("PassengerId")
```



```
clear missingIndices2
```